

REQUEST FOR EXPRESSION OF INTEREST (REOI)

Financial consumer protection suite for supervisory agencies with web scraper and ML-based analysis

Project: Development of a working prototype for a financial consumer protection suite with web scraper and AI-analysis (the "Project") serving two financial authorities in one country (the "Agencies")

Description: An advanced analytics tool capable of processing streams of web and social media channels, and linking with entities and trends in historical complaints data, using machine learning (ML) to provide a comprehensive view of the market in the interest of financial consumer protection supervision.

Contracting Entity: University of Cambridge, Judge Business School

Countries and Agencies: TBC before vendor selection and contracting

Grant Value: US\$125,000

Publication Date: 24 February 2023

Expression of Interest Deadline: 09 March 2023 23:59 GMT Time (UTC +0)

Project Implementation Dates: May 2023 – November 2023

Procurement Process Managed by: Cambridge SupTech Lab at the Centre for Alternative Finance, the University of Cambridge Judge Business School

Submission: Email all documents to the Cambridge SupTech Lab's Launchpad at suptech-launchpad@jbs.cam.ac.uk with subject line "REOI: Cambridge SupTech Lab – AI-Powered Consumer Protection Suite Project"

Language: All submissions must be written in English.

The Cambridge SupTech Lab

The Cambridge SupTech Lab ("the Lab") at the Cambridge Centre for Alternative Finance, the University of Cambridge Judge Business School, accelerates the digital transformation of financial supervision.

While financial services are becoming increasingly global, digital and complex, analogue processing and antiquated technologies in data gathering, validation, storage and analysis erode the analytical capabilities of supervisory agencies, who are often too late in protecting consumers from fraud and seeing signs of stress in the financial system, or miss the underlying causes. This is all happening while financial crime remains a trillion-dollar issue, and public agencies face new challenges such as the regulation and supervision of crypto assets, and monitoring environmental, social and governance (ESG) aspects of the financial industry's business.

The Lab aims to meet financial sector supervisors' needs by developing with them new methodologies and processes that further market oversight and empower consumers, and to deploy suptech applications that generate relevant, reliable, timely insights to inform their decisions. From research to executive education, to technical assistance, to crafting production-grade suptech solutions, we are committed to supporting the emergence and acceleration of the suptech ecosystem and to empowering a new generation of innovation leaders seeking to digitally transform financial supervision.

For more information about the Lab, please visit <https://www.cambridgesuptechlab.org>

SupTech Launchpad

To accelerate the growth of the suptech marketplace, the Lab partners with financial authorities and technology vendors to co-create and deploy cutting-edge, scalable suptech applications. Our team helps detail the technical specifications, de-risk procurement for all parties, and provide project management support and hands technical assistance including security testing. Furthermore, we provide the vendors with coaching and opportunities to engage with investors and other stakeholders.

The most transformative Proofs of Concept (POCs) developed through the [Innovation Leadership Programme](#) are selected for agile prototyping and deployment. The process involves global competitions to crowdsource ideas from technologists and identify the best implementation partners among both other financial authorities and vendors. The Lab's Launchpad largely builds on the experience of the RegTech for Regulators Accelerator (R²A), which successfully developed [groundbreaking applications](#) by introducing an agile mode of collaboration between financial authorities and vendors.

For more information, please visit <https://lab.ccac.io/launchpad/>

I. Project Description

The focus of the project is to develop an AI-Powered financial consumer protection suite, which is comprised of two primary components:

1. **Web Scraper and AI Analysis:** the facilitation of web scraping of open sources (i.e. public sites, social media, app store), sources on keywords and terms that AI can expand or reduce using sentiment analysis based on continuously scraped data that correlates with closed sources (i.e. current chatbot solution, business registry), in order to facilitate early warning of possible questionable business practices, scamming, and fraud. Identified companies and persons can be fed into the Agencies' case management tools.
2. **Data Management System:** development of an integrated data management system (DMS), and integration with complaints data sources (e.g., traditional support channels, plus an existing chatbot solution), subsequent analysis, and development of relevant reports based on these data.

This solution will add additional web and social media channels for consumer protection data collection and analysis, as well as to incorporate the ability for supervisors to analyze these against other sources of explicit complaints raised prior chatbot deployments from the Agencies.

Value by stakeholder type

The Agencies will be able to:

- Incorporate open/alternative sources of monitoring information that are currently being monitored manually
- Cross-reference monitoring information with current commission company registries and complaints/service request ticketing systems
- Automate the incorporation and cross-referencing processes previously indicated
- Use artificial intelligence (AI) to be able to handle larger data sources and produce both descriptive and predictive results leading to more-proactive monitoring
- Provide such monitoring services across all departments, with an emphasis on regulating and enforcement activities
- Manage cases more efficiently and proactively.

The Agencies' supervisors will be able to:

- Access more comprehensive information sources not dependent on companies and their voluntary submissions
- Benefit from AI that ensures efficient and more selective filtering
- Automate existing manual processes, leading to less effort and time saved

- Deploy an early warning system enabled by the predictive nature of AI, allowing for proactivity
- Cross reference open and closed information sources.

Regulated firms will be able to:

- Have partial relief from perceived costs of compliance while having access to new sources of information that provides added value
- Access feedback mechanism that allow for better market response
- Have quantified complaints statistics that can provide insights to customer expectations and better means to measure overall market satisfaction.

All users (potentially including the general public) will be able to:

- Have access to metrics that are produced on a continuous basis and that can be a measure of trustworthiness for investors or borrowers
- Have access to service providers with less risks that can rise to the higher consumer satisfaction levels for better consumer choice
- Lodge complaints that can have positive effects for providers as well as consumers.

II. Description of Required Solution

Basic requirements

The proposed suptech solution has the following components:

1. Web-scraping Tool: the component features of the Web-scraping tool are:
 - i. Configuration Interface: Investigators must have this tool/interface to input entries such as watch list details and resource details
 - ii. Crawler Tool: Investigators must have this tool to automate tedious, inefficient, incomplete, and time-consuming tasks
 - iii. Collection Tool: This tool could provide interfacing to complaint and service request ticketing tools submitted directly to the Agencies
 - iv. Correlation Tool: Investigators must have this tool to correlate crawled web-scraping results to existing company information submitted to the Agencies.
2. Case Monitoring Tool: the component features of this tool are:
 - i. Data Filter Tool: Data must be filtered for significance to ensure that only significant and non-redundant information are passed to heavier processing and resource-intensive tasks as well as storage

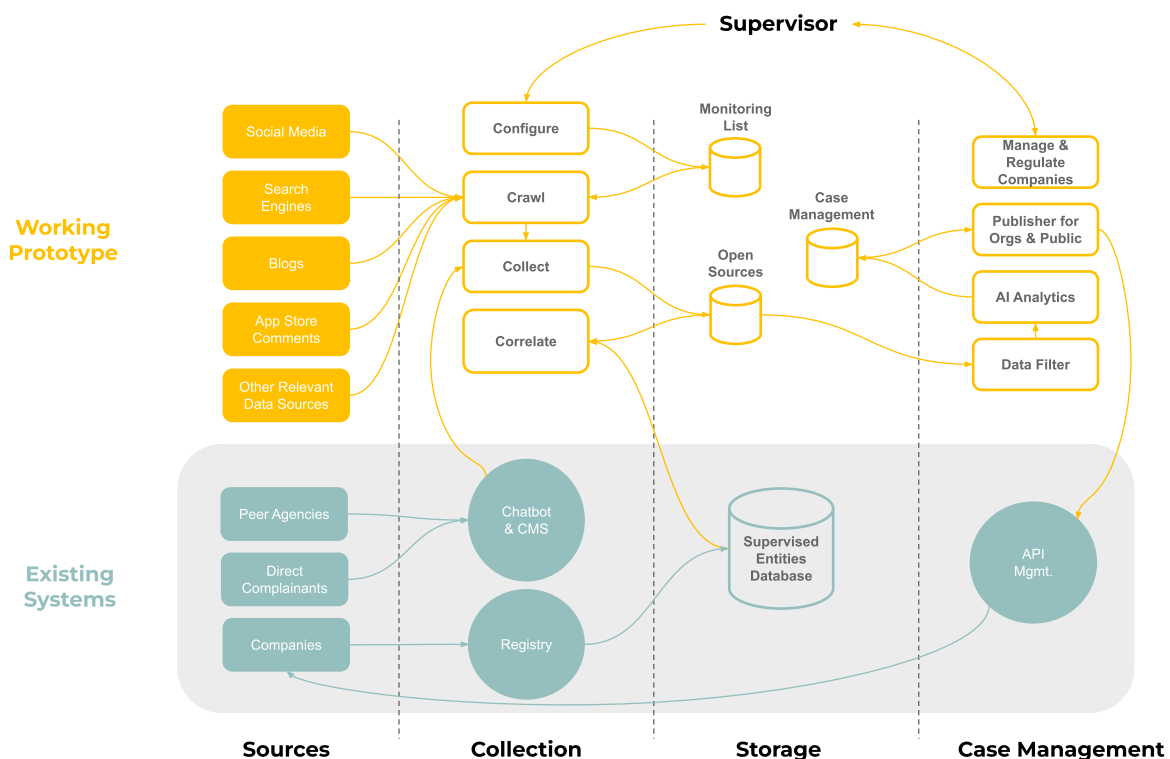
- ii. AI Analytics Tool: Analysis must be performed on filtered data to properly classify and organize data into actionable cases
- iii. Publishing Interface: Unverified data can be endorsed or forwarded to partner enforcement agencies, regulated entities, and the public for possible early warning
- iv. Management and Enforcement Tool/Interface: Investigators must have this tool/interface to view, verify, and tune data and actual cases for enforcement and be able to forward notifications to identified registered companies.

The valid information sources that are covered by this project are:

- Closed data sources: Information resources actively being maintained by the Agencies in pursuance of its mandates including company registration, compliance submission, and monitoring systems, whether planned or existing within the Agencies.
- Open data sources: Information resources external to the Agencies that are either open and freely available or proprietary but whose access the Agencies has lawfully procured such as search engines, social media sites, blogs/vlogs, app store comments, and other sources the Agencies deems necessary and lawful to monitor in pursuit of its mandates.

High-level architecture

This diagram illustrates the relationship between the components and data sources listed above. Existing infrastructure is illustrated in green; the scope for the Project is illustrated in yellow.



Key technical requirements

The following are common components that are universal across all document management system (DMS) software¹:

- Storage engine
- Query language
- Query processor
- Optimization engine
- Metadata catalog
- Log manager
- Reporting and monitoring tools
- Data utilities.

Scope limitations

In Scope: The scope of this project is limited to the piloting of the Integrated AI-Powered Consumer Protection Suite, limited to the following stakeholders:

- Agencies' supervisors and researchers currently engaged in manually monitoring open sources in the interest of building cases against malicious parties engaged in the Agencies' relevant supervisory activities within the jurisdiction.
- Regulated firms (those companies that are licensed by the Agencies within the jurisdiction).
- Transacting public (individuals who are considered investor, consumer, complainant, and borrower customers of the Agencies).
- Other government agencies (peer agencies within the national government who are authorized to transact with the Agencies).

Out of Scope: Development of the currently existing infrastructure, including, for example, the chatbot, the company registry, supervised entities database, and API management system.

III. Key vendor requirements

The Project requires a vendor with the capacity, relevant experience, and resources to design, develop, test, and deploy a prototype with the purpose of web scraping of open sources, DSM, analysis, and integration with current chatbot solution.

¹ <https://www.bmc.com/blogs/dbms-database-management-systems/>

General Launchpad qualifications

- Specificity: the competition is result oriented and the proposed solution needs to have a high level of detail and granularity with respect to the expected output.
- Precedent: the applicant needs to work on a novel solution.
- Geography: the vendor can be based in any jurisdiction. Data needs to be stored in an infrastructure compliant with the needs of the financial authority.
- Collaboration: the development of the solution should be conducted in collaboration with the team designated by the financial authority in each distinct phase.
- Sufficient experience to build an application that can serve the data needs of the financial authority, e.g.:
 - Integrate with existing application within a governmental entity
 - Allow the migration of existing data
 - Provide real time, on-demand support and the ability to generate reports or summary
 - Provide a high standard of application security
- Demonstrated ability to:
 - Manage product life cycle
 - Develop, complete, implement, maintain, and deliver the appropriate technologies
 - Properly write documentation
 - Maintain an enterprise ecosystem.
- Experience with:
 - Suptech solutions
 - Working with regulators and financial authorities
 - Frontend Development and UI/UX Design (as needed)
 - Data products and practical applications of analytics and data science (e.g., AI/ML)
 - Software engineering
 - Application architecture and devops
 - Program/project management and business analysis
 - Agile methodologies for application development
 - Application integration and performance tuning/optimization

- Knowledge including:
 - Cybersecurity and secure application development coding standards
 - Best practices in relevant fields to the solution at hand
- Resources:
 - Technical expertise on related knowledge and experience
 - Sufficient staffing and computing resources required by the identified feature and time requirements and constraints
 - Sufficient specification of online, on-premises, and/or hybrid computing resources
 - Adequate project management staffing based on requirements
 - Software, hardware, network, and cloud computing licenses and subscriptions to cover development, implementation, and warranty period
- Project management:
 - Methods that leverage agile delivery methodologies for project planning, design, building and testing, stakeholder engagement, and effective risk management to ensure on-time completion of the project without budget overrun.

Engagement-specific requirements

- i. Capability:
 - Develop, complete, implement, maintain, and turnover Web scraping, Business Intelligence, Artificial Intelligence, and Machine Learning Full Stack solutions.
 - Implement Full Stack solutions that span data components, business logic, automated and scheduled processing, orchestration, and user interface development and implementation.
- ii. Experience:
 - Web scraping of unstructured public data sources
 - Processing of online textual information using world-class NLP technologies at scale, for a combined public, industry, and regulatory audience
 - Experience designing and developing integration specifications for public sector chatbot systems
 - Frontend development and UI/UX design for diverse user basis of all levels of technical prowess
 - Deploying additional technical capacities at scale, including:
 - DBA/Big Data

- Software engineering
- Data, Systems, Technical, and Application Architecture
- Program/Project Management and Business Analysis
- Data Analytics
- Artificial Intelligence/Machine Learning.

iii. Knowledge:

- Agile Method for Application Development
- Full Stack Development
 - Data Models
 - Backend Business Logic
 - Orchestration
 - Front End.
- Data Mining/Business Intelligence and Artificial Intelligence/Machine Learning
- Advanced Data Collection Techniques
- Application Integration and Performance Tuning/Optimization
- Cybersecurity and Secure Application Development Coding Standards
- Program/Project Management
- Product Management
- Cloud Computing

IV. Project Award

The successful applicant will:

- Be awarded **US\$125,000** to develop and test the required solution with two financial authorities in the same country. This is a fixed-sum contract, which is to cover all the applicant's expenses related to the development and testing work, including staff time, hardware, software, travel, and all other project-related expenses.
- Receive tailored coaching.
- Be invited to the Lab's pitch day to connect with funders and to a demo day to present their products to potential clients.
- Be listed in the Lab's online [Vendor Database](#).

- Be mentioned in a case study published by the University of Cambridge to share lessons from the project.
- Engage with the suptech community through the Lab's hackathons and techsprints.
- Be introduced to the global community of regulators and supervisors, investors, academics and development partners that are collaborating with the Cambridge SupTech Lab during other events hosted by the Cambridge Centre for Alternative Finance (CCAF).

V. Vendor Selection and Project Implementation

Timeline

Following the receipt of Expression of Interest submissions from qualified vendors, a Request for Proposals (RFP) will be formally issued to three shortlisted vendors on March 19th, 2023. The invited vendors will have ten days to submit their proposal. The winner of the RFP will be announced on May 2nd, 2023, with work commencing within three weeks. This Project will ultimately deliver a prototype that will be tested by the Agencies no later than November 2023.

Key features of this initiative

- Blind review process: A panel of expert reviewers will score anonymised proposals without knowing the name of the vendor submitting them.
- Competitor scorecard: Applications will be assessed by the panel using a set of scoring metrics and weighing the relative importance of each attribute.
- Rapid turnaround time: We will select the winning vendor and award 50% of the **US\$125,000** within 48 days from submission of the final proposal. The last 50% of the award will be granted in one installment upon completion of the deliverables according to projected timelines.

Project Structure

The Project has four phases, elaborated below:

1. Kickoff and interface design, including technical integration specifications
2. Development of a working prototype
3. Integration and development of additional data analytics and/or visualization.
4. Testing and signoff of the working prototype

Throughout all project stages, vendors are expected to meet weekly with key stakeholders of the Agencies as well as the Lab's Launchpad team, to ensure close coordination and agility.

1. KICKOFF & INTERFACE DESIGN, INCLUDING TECHNICAL INTEGRATION SPECIFICATIONS

During the first phase, the selected vendor (in coordination with the Lab's Launchpad team) will gather requirements from the Agencies and produce an initial Design Document that includes integration and user interface specifications. This living document should include specifications for the client-facing portion of the prototype, communicated in a manner such that clients of the prototype can understand how to integrate with systems and processes, submit data, and use the system without necessarily understanding the entire architecture behind the software. This includes specifications for the data integration (analytics and visualization) phase of the project as well. This Document is to be shared as needed with any other key stakeholders (e.g., vendors of relevant software used by the Agencies, any financial institutions needing to integrate) to allow them to develop integrations and/or adapt existing software during the development phase. The design document should also include criteria for user acceptance testing for use during the testing phase.

2. DEVELOPMENT OF A WORKING PROTOTYPE

The selected vendor will build a working prototype that delivers on the requirements laid out in the REOI, RFP, Project Agreement, and any agreed-upon modifications to scope agreed during the previous Phase 1.

The AI-Powered Consumer Protection Suite will first receive a small subset (a representative sample) of all data required in a controlled environment. Initial data being submitted via the Suite will be sample data to start, with real data only being introduced to the system once proper security protocols and data sharing agreements are in place. Starting with a prototype and a small data set will allow the vendor to quickly identify and address any unforeseen issues early in the Project development cycle.

Once agreements are in place, the Suite's model(s) can be trained, tested, and validated on real data.

The working prototype will also facilitate candid discussions among Project stakeholders regarding issues such as model interpretability, potential externalities, and the like.

3. INTEGRATION AND DEVELOPMENT OF ADDITIONAL DATA ANALYTICS AND/OR VISUALIZATION

Once the working prototype has been developed, tested, and accepted, the vendor will provide any analytics and visualization tools defined during the design stage.

Additionally, the selected vendor should (i) assess the needs of the Agencies to understand which dashboards, reports, and statistics are most useful and/or difficult to produce under the current system; and then (ii) propose and develop a prototype mechanism for extracting and visualizing this information from data consumed, processed, and produced by the Suite working prototype. This could involve creating custom queries, scheduling the generation of reports, and outputting in various formats.

The final UI of the prototype must be done before the final tests of Phase 4.

4. TESTING AND SIGNOFF OF THE WORKING PROTOTYPE.

Once the proof of concept has been completed to the satisfaction of the involved parties, integration and testing with real institutional data can begin. The working prototype will be tested with the Agencies, based on any user acceptance criteria defined during the design stage, to allow the vendor to ensure that the prototype Detector can handle the acceptance, reporting, and detection of fraud within real data before it is scaled into full production. This approach also minimizes the risk of interruption due to unforeseen technology failure and serves to inform estimates of the cost to scale the prototype to a production-grade service.

A cyclical final test of the prototype and improvements by the developers must be done until the product is adherent to the functional specification document.

VI. Rules and guidelines

Submission Requirements

Interested vendors must demonstrate that they are qualified to perform the services required for the Project. In particular, interested vendors are asked to address the following requirements in their submission in a format capable of being read by Microsoft Word, which should together be **no more than 8 pages in length** (minimum 11-point font):

1. Company background (including technical and managerial capabilities of the executives).
2. A list of past projects representative of the experience of the company and the executives.
3. Information on the qualifications of key staff to be involved in the Project, including whether they have experience working on supotech projects.
4. Summary of your working experience in the following geographies: (1) Philippines; (2) Indonesia; (3) Ghana; (4) Peru; (5) India.
5. Examples of prior experience related to the development of this solution prototype or similar technological solutions.
6. Indicative development / implementation schedule based on the timeline set out above.

7. Detailed description and examples of technical prowess that address both components of the key vendor requirements in section III.

All materials should highlight the relevance to the required solution described in this document.

Supporting documents may be submitted (e.g., company brochures, case studies, CVs, etc.) as attachments to the submission email. No page limits apply to these attachments, but evaluation will be based primarily on information included in the main body of the EOI.

Any questions prior to the submission deadline should be sent via email to Cambridge SupTech Lab's Launchpad at suptech-launchpad@jbs.cam.ac.uk with subject line "REOI: Cambridge SupTech Lab – AI-Powered Consumer Protection Suite Project"

Disclaimer

This document is not a request for proposals. The Lab reserves the right to edit, invalidate, terminate, and/or reissue this REOI at any time and for any reason. The Lab also reserves the right to select a vendor through an alternate method and/or adopt an alternate timeline for vendor selection that differs from the method and/or timeline described in this document, the websites of the Lab and Launchpad, and any other communications related with the process. Furthermore, the Lab expressly disclaims responsibility for any costs incurred by any vendor in responding to this REOI, regardless of whether the REOI is edited, invalidated, terminated, and/or reissued at any time and for any reason.